

Geospatial Statistics on Big Data: Scalability and Applications

Amr Magdy

University of California Riverside
Department of Computer Science and Engineering &
UCR Center for Geospatial Sciences

Statistics beats the random chance

- > In a workshop, are there teenagers (ages 13-19)?
 - > 1 sample person is random
 - > 30 sample persons are statistically significant (better confidence)



(Geo)Spatial Statistics

vs Traditional Statistics



In a spatial space, iid assumption is invalid

independent and identically distributed

The ages of attendees are independent, and each person has the same weight in the sample (identical)



(Geo)Spatial Statistics

vs Traditional Statistics



In a spatial space, iid assumption is invalid

independent and identically distributed

Al Shamal
Temperature: ?

Al Rayyan
Temperature: ?



Doha Temperature:
33 C

Spatial Autocorrelation

Applications of Geostatistics

› Urban Segregation

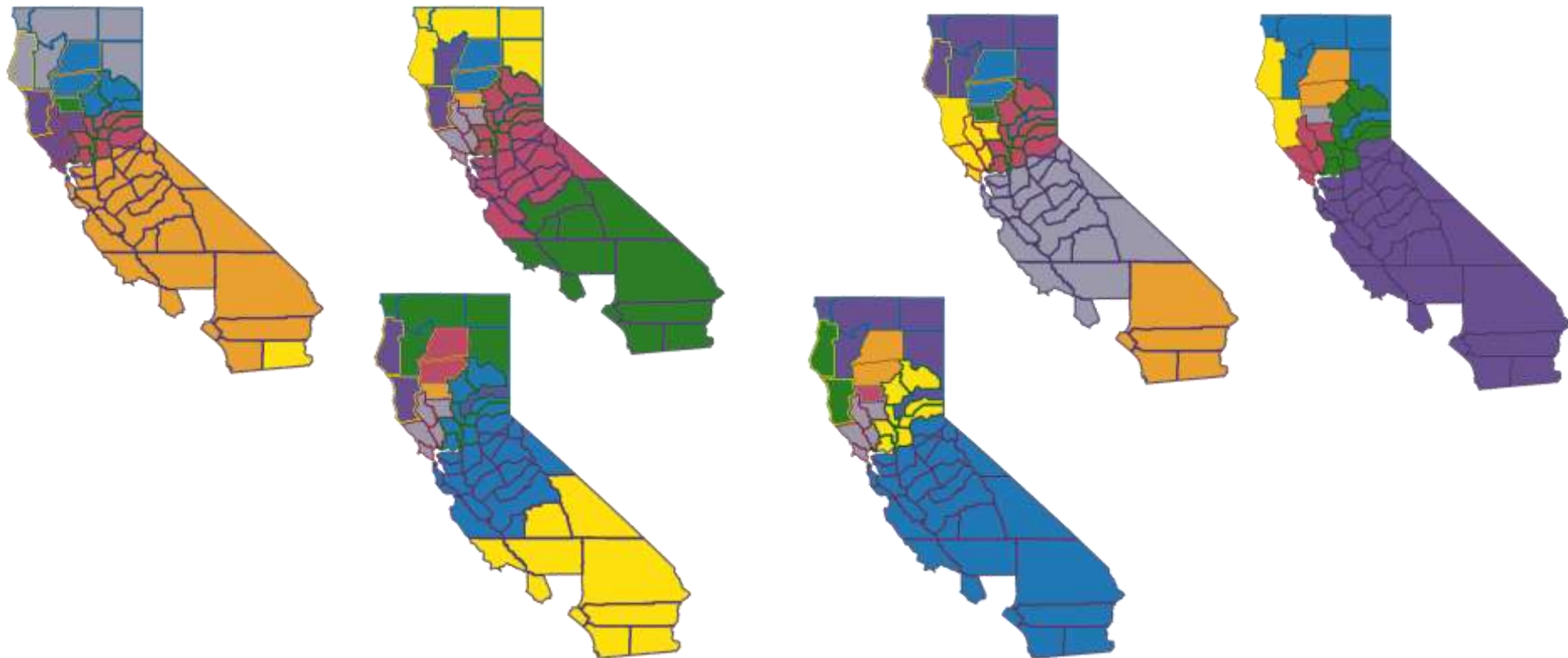
- › Divisions of the urban space based on: ethnicity, income, culture (language/dialects), etc
- › Example: Black and Hispanic segregation in the USA



Applications of Geostatistics

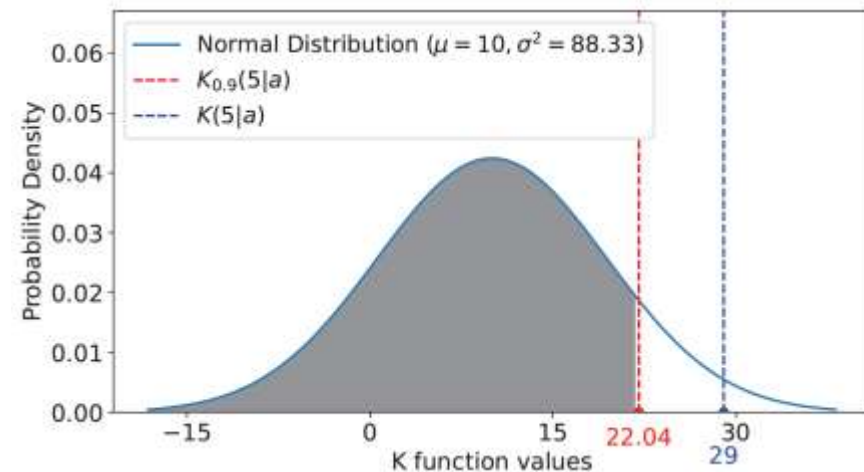
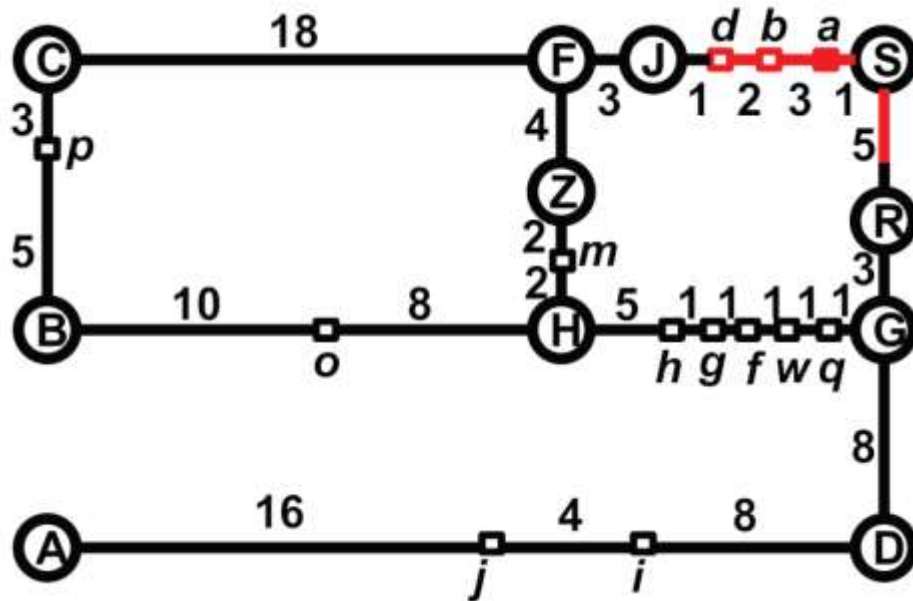
› Urban Segregation

- › Divisions of the urban space based on: ethnicity, income, culture (language/dialects), etc
- › Example: Black and Hispanic segregation in the USA



Applications of Geostatistics

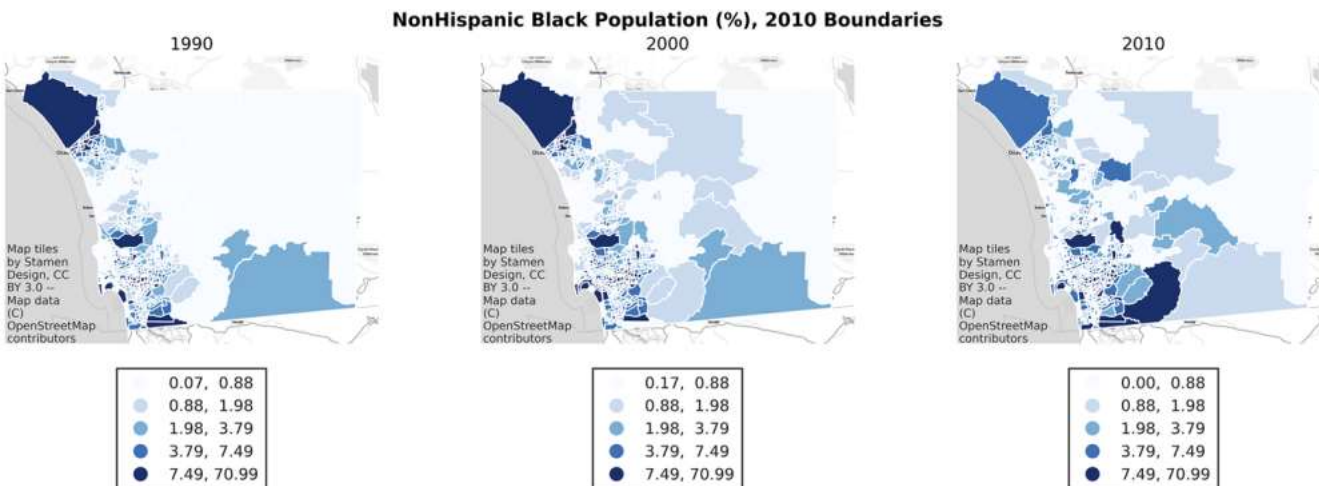
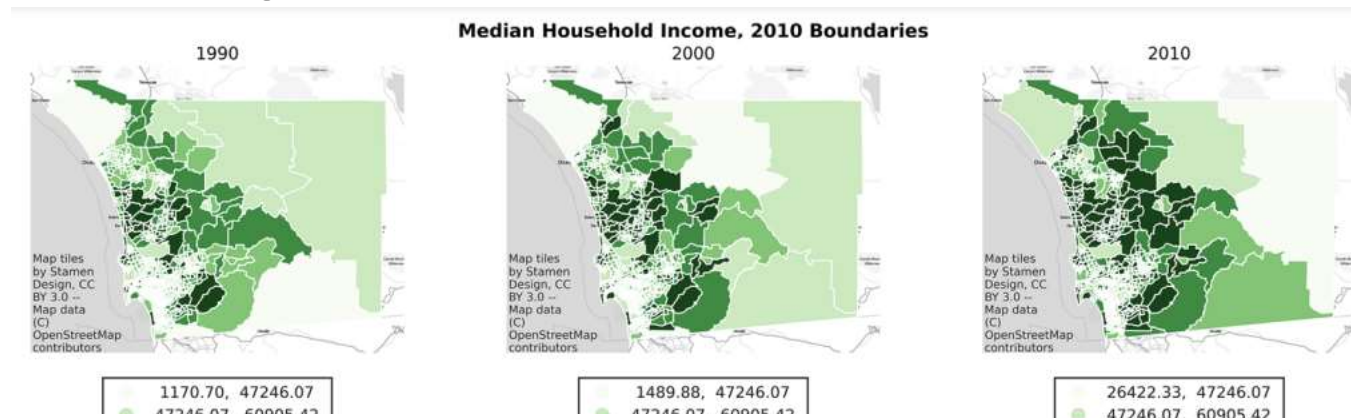
- > **Statistically-significant hotspots**
 - > Finding regions of concentrations with statistical significance
 - > Examples: crimes, traffic accidents, disease outbreaks



Applications of Geostatistics

› Re-defining / Analyzing Neighborhoods

› Example: economic growth, inequality, education quality, etc



Big Data Role

(1) Data Preparation

- Examples:

- Grouping lines (road segments) into polygons (city blocks)

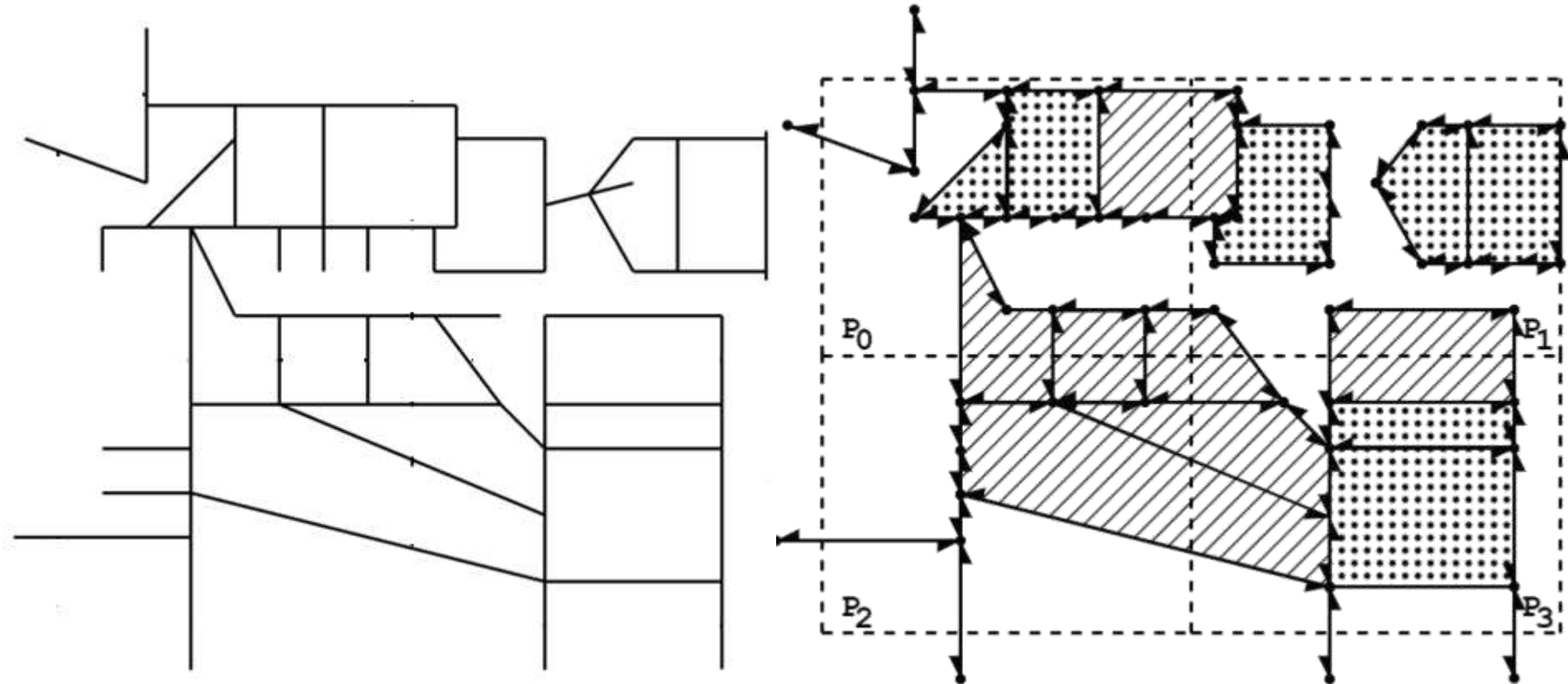
- Grouping polygons (blocks) into regions (neighborhoods)

(2) Data Processing

Big Data Role

(1) Data Preparation

- Examples:
 - Grouping lines (road segments) into polygons (city blocks)



Big Data Role

(1) Data Preparation

- Examples:
 - Grouping lines (road segments) into polygons (city blocks)

Dataset	Area	Size
USA	9.83 Mkm^2	152M
South America	17.8 Mkm^2	155M
North America	24.7 Mkm^2	240M
Africa	30.4 Mkm^2	288M
Europe	10.2 Mkm^2	563M
Asia	44.6 Mkm^2	557M

Big Data Role

(1) Data Preparation

- Examples:
 - Grouping lines (road segments) into polygons (city blocks)
 - Grouping polygons (blocks) into regions (neighborhoods)

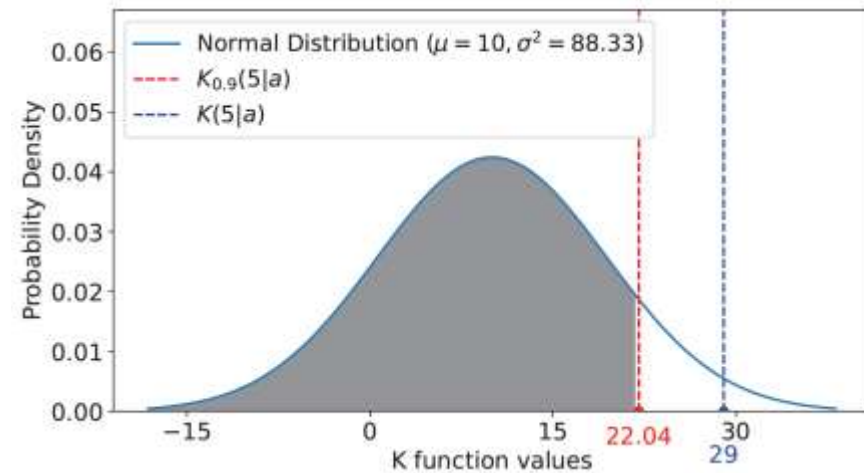
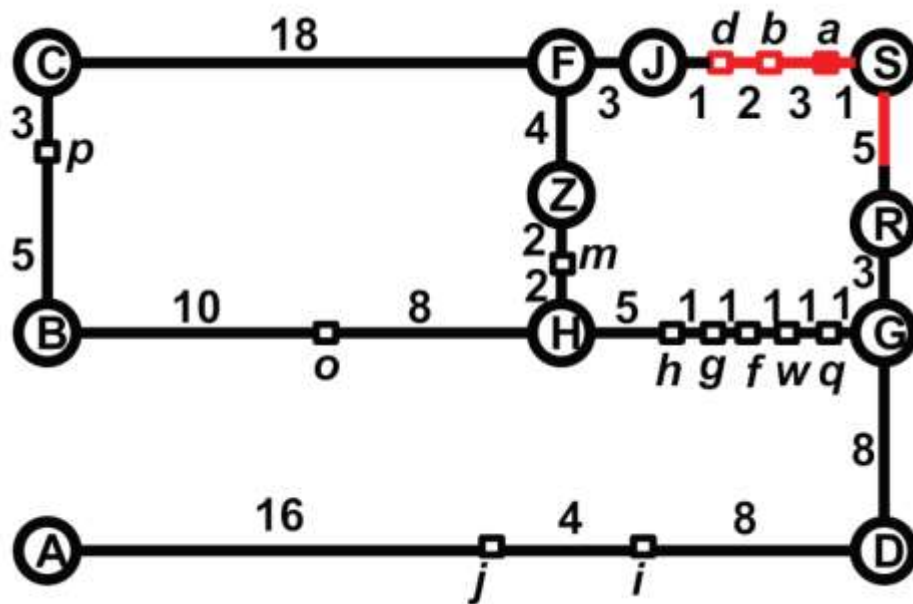
(2) Data Processing

- Scaling up statistics computation
- Enabling rich user-defined constraints
- Examples: detecting hotspots from millions of points

Big Data Role

(2) Data Processing

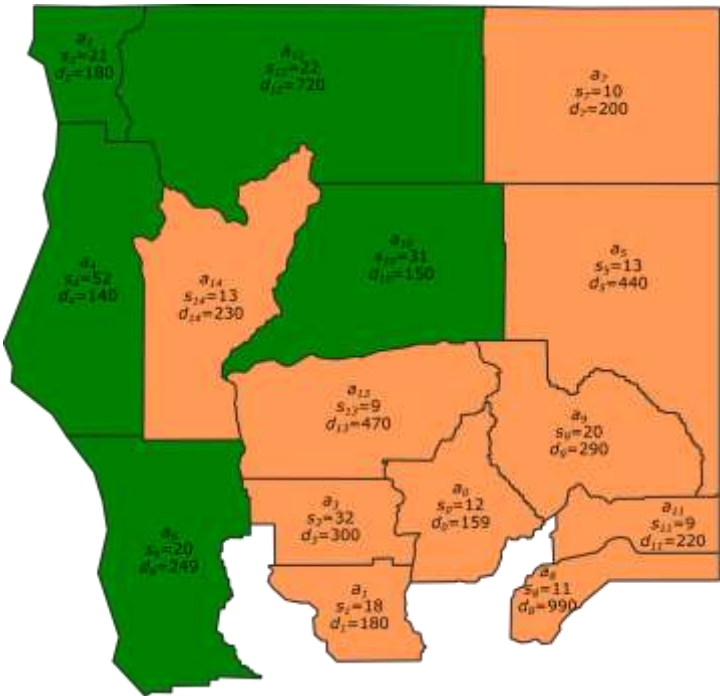
- Scaling up statistics computation



Big Data Role

(2) Data Processing

- Enabling rich user-defined constraints



Discover regions with:

- * **Total** population \geq 200K,
- * $\$1K < \mathbf{average}$ income $<$ $\$3K$, and
- * $10K < \mathbf{Min}$ number of COVID infections $<$ 20K

Take-home Messages

- › Data in spatial spaces are correlated
- › Geostatistics provides confidence-based spatial data analysis
- › Applications span many societal applications in studying, monitoring and addressing social issues
- › Big data technology helps to analyze larger regions and provide more sophisticated analysis

Thank You

Questions?

<http://cs.ucr.edu/~amr>

Email: amr@cs.ucr.edu